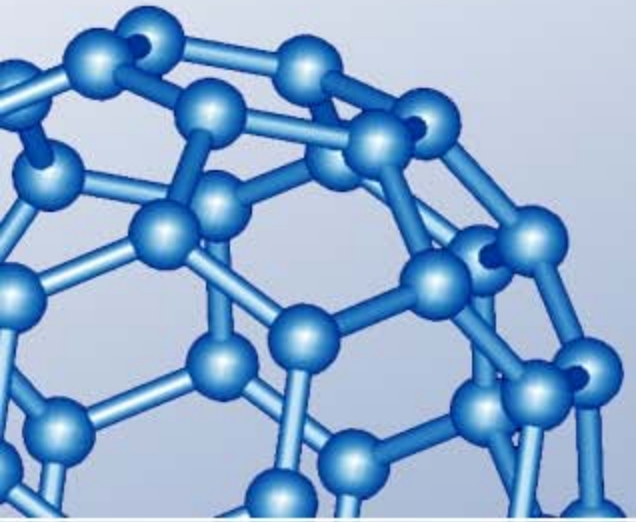
Nanoinformatics in Europe: The ACTION-Grid White Paper

Victor Maojo, Diana de la Iglesia, Miguel García Remesal; Fernando Martin-Sanchez, José Crespo, George Potamias, Vassilis Moustakis, Paula Otero, Sonia Benitez, Fernando Gonzalez-Nilo, Yannick Legre, Josipa Kern, Julio Facelli, Joyce A. Mitchell, Casimir Kulikowski



Victor Maojo, MD, PhD
Director, Biomedical Informatics Group,
Universidad Politecnica de Madrid (UPM)
vmaojo@fi.upm.es

Arlington, VA, 3-5/11/2010

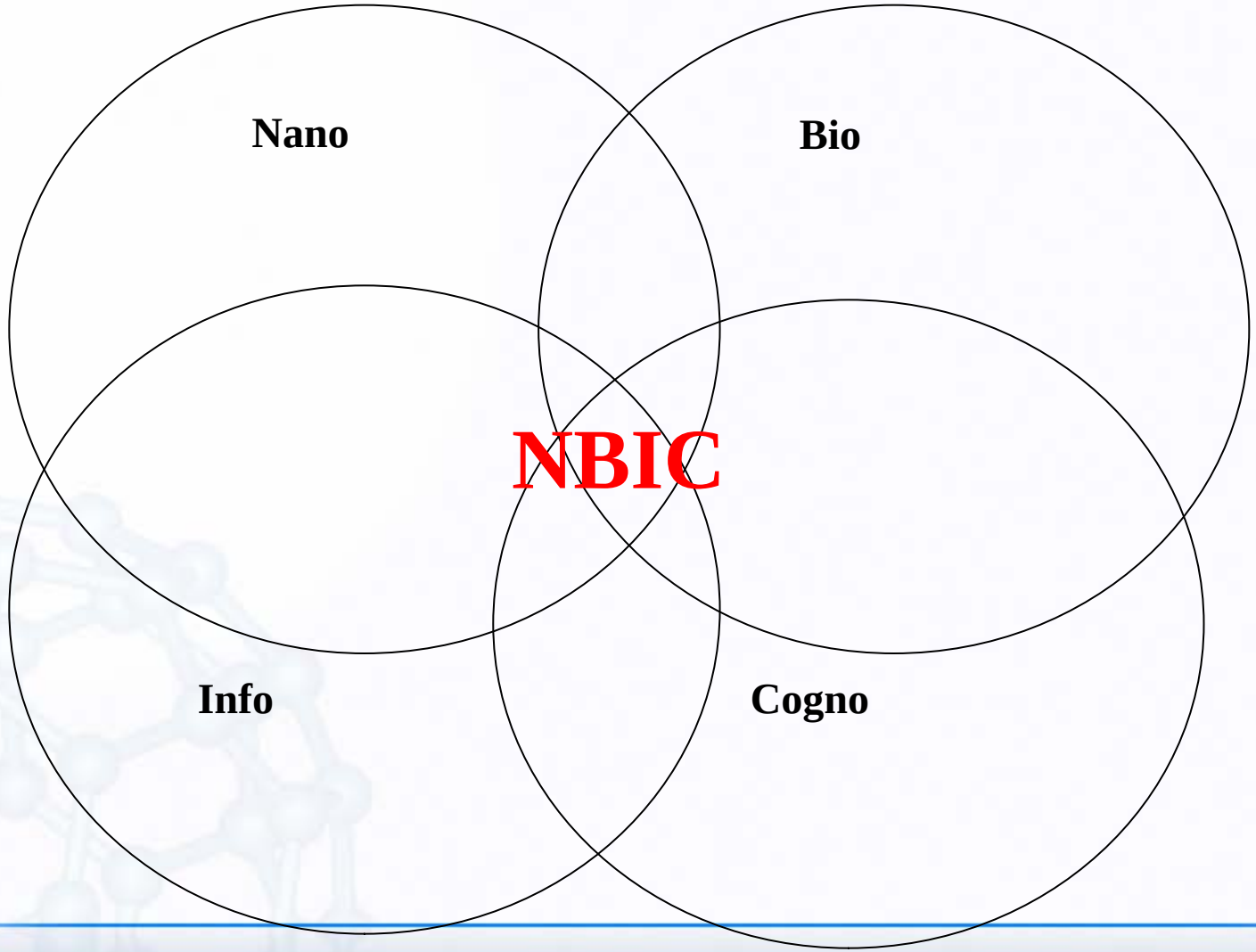
The Action-Grid project was funded by the European Commission under the 7th framework program

European Commission's Framework Programme 7: Virtual Physiological Human

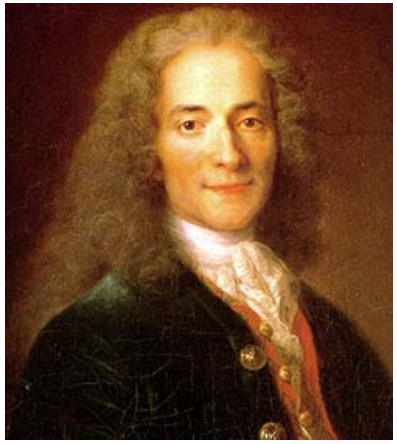
- Centered on building models and simulations of the body linked with clinical data
- A long-term, challenging vision
- ACTION Grid: only project related to BMI and nano areas within its Call
- Challenge: expand the VPH scope with a "nano" perspective

The Action-Grid project is funded by the European Commission under the FP7

Expanding the scope towards Converging Technologies



The great philosopher Pangloss, 1st pioneer in Converging technologies!!



From Candide (Voltaire, 1759)

"Master Pangloss taught the ***metaphysico-theologo-cosmolonigology***. He could prove to admiration that there is no effect without a cause; and, that in this best of all possible worlds, the Baron's castle was the most magnificent of all castles, and My Lady the best of all possible baronesses".

Ubiquity: Technologies for Better Health in Aging Societies
A. Hasman et al. (Eds.)
IOS Press, 2006

Reflections on Biomedical Informatics: From Cybernetics to Genomic Medicine and Nanomedicine

Victor MAOJO[a] and Casimir A. KULIKOWSKI[b]
[a]*Biomedical Informatics Group, Artificial Intelligence Lab, Universidad Politecnica de Madrid, Spain*
[b]*Dept of Computer Science, Rutgers Univ. New Brunswick, New Jersey, USA*

Abstract. Expanding on our previous analysis of Biomedical Informatics (BMI), the present perspective ranges from cybernetics to nanomedicine, based on its scientific, historical, philosophical, theoretical, experimental, and technological aspects as they affect systems developments, simulation and modeling, education, and the impact on healthcare. We then suggest that BMI is still searching for strong basic scientific principles around which it can crystallize. As -omic biological knowledge increasingly impacts the future of medicine, ubiquitous computing and informatics become even more essential, not only for the technological infrastructure, but as a part of the scientific enterprise itself. The Virtual Physiological Human and investigations into nanomedicine will surely produce yet more unpredictable opportunities, leading to significant changes in biomedical research and practice. As a discipline involved in making such advances possible, BMI is likely to need to re-define itself and extend its research horizons to meet the new challenges.
Keywords: Biomedical Informatics, Medical Informatics, Bioinformatics, Ontologies, Nanomedicine, Virtual Physiological Human

1. Introduction

Several analyses of the current scientific status of Medical Informatics (MI), Bioinformatics (BI), and Biomedical Informatics (BMI), that has evolved over the last few years from the convergence between MI and BI, have been published over the last decade [1][2]. The authors of this paper have earlier contributed to this debate [3][4].

MI researchers have sometimes stated they encounter difficulties in receiving the scientific recognition that they think the discipline deserves [4]. In some cases, MI has been considered a pure engineering discipline, focusing on the technology related to health care. Could this situation be only circumstantial or might it reflect a trend that may become further accentuated? What could be the future of MI and BI? Will they merge into BMI or will they remain independent?

2. Proposal for a more systematic analysis of MI, BI, and BMI

In a series of various papers, the authors have analysed the scientific status of these three disciplines [3][4], using as a framework various ideas from the philosophy of science. We extend these ideas below, following various directions.

2.1. Historical

One of the goals of cybernetics was to study the components of what can be referred to as "human information systems". In contrast, MI has not made this a main concern.

MIE 2006:
Maojo and Kulikowski: Reflections on Biomedical Informatics: From Cybernetics to Genomic Medicine and Nanomedicine (First reference on Medline combining “Biomedical Informatics” and “Nanomedicine”)

The Action-Grid project is funded by the European Commission under the FP7

Nanoinformatics: beginning (2007)

Workshop on Nanoinformatics Strategies

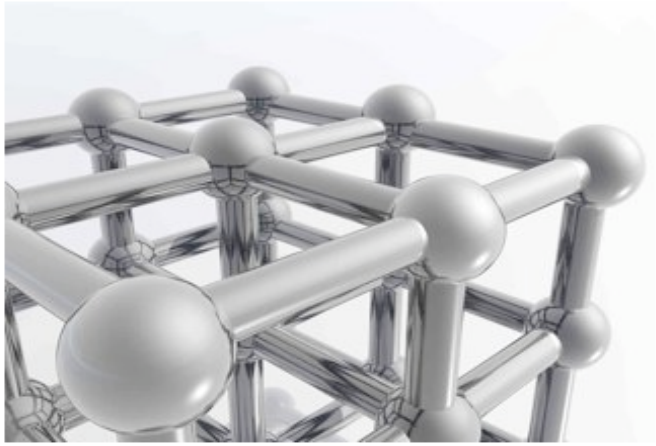
June 12-13, 2007, Westin Gateway Hotel, Arlington Virginia

Hosted by the National Nanomanufacturing Network

Agenda (talks & links)

Workshop Purpose

Participants



National Nanomanufacturing Network

The Workshop on Nanoinformatics Strategies was supported by the National Science Foundation through a grant to the NSF Center for Hierarchical Manufacturing at the University of Massachusetts Amherst.

Workshop goals (Tuominen and Kim, 2007, before ACTION Grid)

- Identify nanoinformatics needs, challenges and priorities
- Discuss informatics activities currently underway that work to address needs in various research, development and education sectors (NCN, EHS, NNN, NIST, NCLT, nanomaterials and others)
- Share best practices on cutting edge techniques in data mining, visual analytics, Web 2.0 technologies, literature analysis, data standards, digital clearinghouses, web-based communication tools, and related topics, including those from the other fields (e.g., caBIG, bioinformatics, computer science and others).
- ***Connection to bioinformatics, and other informatics areas***
- Discuss interconnecting databases and mechanisms for defining the ontology of terms
- Identify and prioritize strategies best suited for catalyzing nanotechnology research, development and education

ACTION Grid Objectives

Main objective

To establish a collaborative environment in Biomedical Informatics, Grid Computing and nanomedicine, among the following geographical areas:

- European Union
- Latin America
- Western Balkans
- North Africa

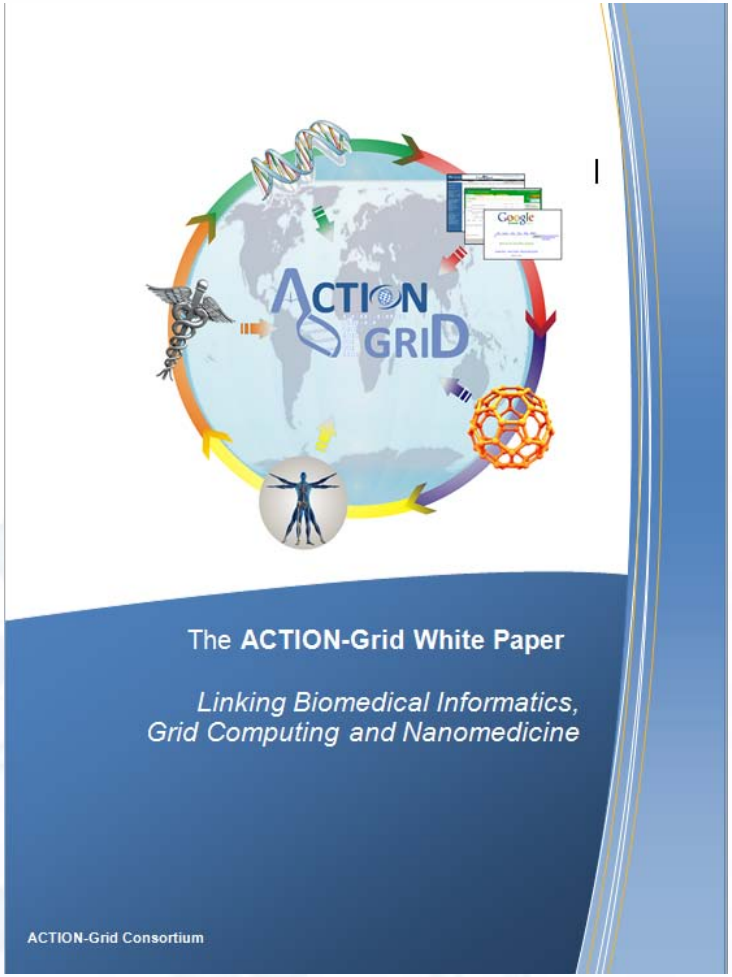
The Action-Grid project is funded by the European Commission under the FP7

Action-Grid: a puzzle with many pieces (some of them not well defined)

The Action-Grid project was funded by the European Commission under the FP7

Victor Maojo

The White Paper: scientific emphasis on Nanoinformatics (freely available at the ACTION Grid website, in large and summary versions)

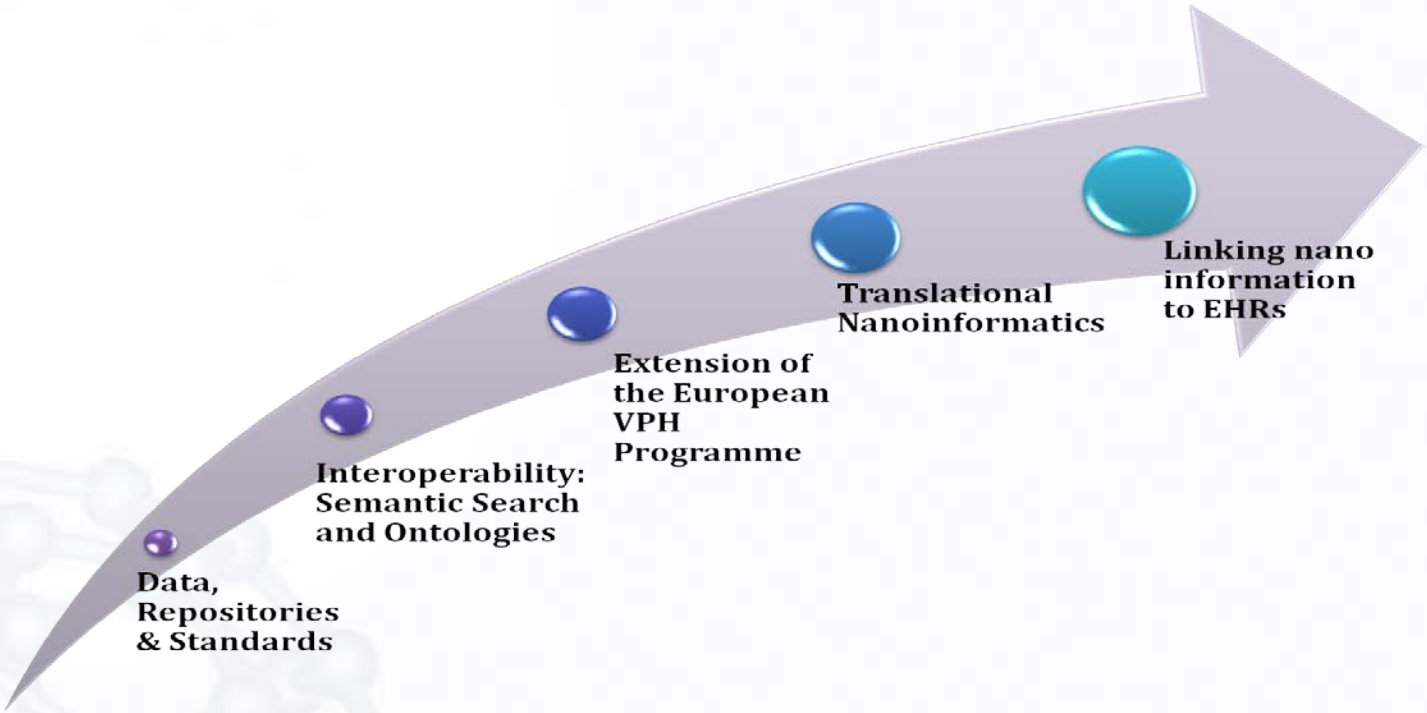


- State of the Art of Nanoinformatics
- Analysis of areas:
  - Nanomedicine
  - Nanotechnology
  - Nanotoxicity
  - Public health issues
- Priorities
- Grand Challenges
- Conclusions

http://www.action-grid.eu/index.php?url=whitepapernano

The Action-Grid project is funded by the European Commission under the FP7

Proposal: Five Grand Nanoinformatics challenges (particularly from a pure informatics perspective)



The Action-Grid project is funded by the European Commission under the FP7

1. Data, repositories & standards

- ***To create a Nanoinformatics infrastructure*** to collect, curate, annotate, organize and archive the available data
- ***Design of extended web nano portals***, linking groups and information around the world to facilitate data sharing
- ***To build repositories/databases of use cases, clinical trials experiments or databases*** with nano-data, facilitating the reuse of the data —like Arrayexpress for genomic data.
- Creating ***a repository of nano-related informatics tools***, which can be accessed and retrieved through specific nano infrastructures

2. Interoperability: semantic search and ontologies

- ***To intercommunicate different nanoportals,*** facilitating rapid exchanges and sharing of data and other resources
- Extended ***standards for interoperability*** in the nanoinformatics field.
- ***Development of a taxonomy of research tasks*** (rather than biomedical resources) in the BMI area
- ***Improved classification approaches***, to create new hierarchies/taxonomies based on actual physical, chemical, clinical, toxic or spatial characteristics, added to pure ontological/semantic information
- Use of ***cloud-computing services and supercomputers to carry out complex computational task***s, such as simulating interactions between nanoparticles and cells of the human body, supporting research in the area of multiscale modelling.
- ***Communication with and between nanosensors and nanodevices*** located in the human body
- To establish ***standards in reporting/publishing results*** in nano-particle/nanotechnology research

Reporting open results from research projects

Science Careers From the journal Science

My Science Career | Find a Job | For Employers | Career Magazine | Grants & Funding | Tools & Tips | Community

Science Home > Science Careers > Science Career Magazine > Science Careers Blog > Open Results from Biomedical Research Projects: Where Are They?

Science Careers Blog

Science Career Magazine

« Older: A Different Kind of High-Throughput Screening

» Newer: USDA's Extramural Research Agency Is Now the National Institute of Food and Agriculture (NIFA)

May 5, 2010

Open Results from Biomedical Research Projects: Where Are They?

The following letter was submitted in response to Chelsea Wald's article Scientists Embrace Openness.

Open Results from Biomedical Research Projects: Where Are They?
Maojo, V., Garcia-Remesal, M, Crespo, J., de la Calle, G., de la Iglesia, D. and Kulikowski, C.

Wald has addressed scientific openness in a recent Science article (1), including data and methods used for research. Advances in software tools for bioinformatics search helps (2), but, just becoming aware of open results of research projects funded by public agencies* — e.g., databases, software, papers, e-books* — and finding them efficiently still proves harder than it should.

In the course of producing an advanced, automatically generated on-line inventory of bioinformatics resources (3), we analyzed results from research projects publicly funded by the European Commission, Spanish agencies, and the National Institutes of Health. We discovered that finding the complete set of available information reported to have been generated by the projects could prove quite elusive. Non-peer-reviewed summary reports were commonplace, but specifics of electronic resources, with Web locations, were frequently not, even when researchers mentioned their existence as being openly available.

To enable searches with sophisticated text mining, publicly-funded projects should provide a minimum information set including titles, authors, funding agency, annotations with concepts from ontologies or controlled vocabularies that characterize the functionalities of the resources, papers reporting significant findings using these resources "peer-reviewed quality indicators," and their Uniform Resource Identifiers (URIs).

About this Blog

Get frequent updates from the science-career trenches including advice, opinion, news, funding opportunities, and links to other career-related resources. Our bloggers include Science Careers editors and staff, and select outsiders, including:

Jim Austin, Editor of Science Careers, @SciCareerEditor on Twitter

Alan Kotok, Managing Editor

Elisabeth Pain, Barcelona-based Contributing Editor

Kate Travis, Contributing Editor in Cambridge, U.K., and the Editor of CTSciNet

José Fernández, Community Manager for CTSciNet and MySciNet

Brianna Blaser, Project Director, Outreach

Donisha Adams, Program Associate for GrantsNet

Our guest bloggers include:

Dan Albert MD, MS, Emmett A. Humble Distinguished Director of the Eye Research Institute at the University of Wisconsin.

Beryl Lieff Benderly, author of the Taken for Granted column on Science Careers;

Siri Carpenter, a regular contributor to Science Careers;

Chelsea Wald, a freelance science writer and contributor to Science Careers;

If you are interested in blogging for Science

ADVERTISEMENT

Science Poster SERIES

Browse and download posters.

Access Now!

Featured Jobs

TENURE-TRACK FACULTY POSITION
Utah State University
Logan-UT-United States

TENURE-TRACK FACULTY POSITION
Williams College
Williamstown-MA-United States

TWO TENURE-TRACK FACULTY POSITIONS
York College of Pennsylvania
York-PA-United States

JUNIOR or SENIOR FACULTY POSITIONS
Yale School of Forestry & Environmental Studies
New Haven-CT-United States

Physician-Scientist Faculty Position in Im...
Columbia University
New York-NY-United States

Assistant Professors of Biology

A proposal for reporting open results from Biomedical Research projects, reported at the Science Careers Bog, from the Science journal

The Action-Grid project is funded by the European Commission under the 7th framework project

Victor Maojo

3. Extending the European VPH programme and US initiatives

- BMI researchers have created a large number of ***models and simulation tools which could be reused*** or adapted to nanomedicine.
- To create a hypothetical, "***extended nanotype***" —including a large catalog of nanoparticles and biological targets, their interactions, potential nanotoxicities and relations to different diagnostic and therapeutic uses
- To ***simulate "in silico" the effects, reactions or toxicity*** of new compounds or materials before the "in vivo" studies. Multilevel simulations might predict effects of nanoparticles
- Theoretical studies of the ***interactions between nanoparticles with the most common components of human cells***

4. Translational nanoinformatics

- Nanoinformatics might imply a continuum of BMI, but new insights are needed for ***data and knowledge integration at the nano level*** as well as basic research.
- The nature of nanomaterials and the ***unknown effect of many nanoparticles*** must be addressed prior to semantic or ontological analyses
- M***edical imaging.*** A key issue for nanomedicine is to create new contrast agents to target specific organs, functions, or cell types.
- We still lack a ***unified theory of biomedical information*** in an area like BMI. The analysis of differences between bits and quantum bits —a measure of information at the quantum level—, their appearance in nature and the diversity of meanings of the term “information”, might lead to new insights and research in this fundamental area

5. Linking nanoinformation to EHRs

- T***o link nanomedicine-related data to patient EHRs.*** New standards will be needed for storing data or augmenting clinical vocabularies and terminologies —like SNOMED— or for exchanging electronic medical information —like HL7— and how they can incorporate nano-related information, terminologies and procedures
- Questions related to ***patient safety*** and possible secondary effects related to the use of nanoparticles need to be addressed
- The creation of large databases that would store nano-related information can be complemented by ***new approaches to building EHRs.*** It will require a collaborative effort from a number of researchers
- New diagnostic and therapeutic methods based on new nanomaterials can enhance recent proposals for "personalized medicine" —currently mostly being built based on "–omics" advances. New models of EHRs, including nano information, must be ***developed for use by medical professionals***,

New industrial areas of development. New informatics techniques may allow new companies and academic settings to work on nanotechnology and nanomedical research, without the large budget resources that are usually necessary for traditional research in the area

New synergy between established disciplines like Biomedical Informatics and Nanomedicine, leading to a new subdiscipline or area which could be called Nanoinformatics

Enormous financial importance, with exponential growth in terms of investment and revenues.
Large expectations for new job opportunities

New academic programmes can be created or classical programmes —e.g., BMI or medicine— can be expanded.
New working groups in societies like IMIA, EFMI or AMIA can address nano-related issues.

European groups can be funded by the US NIH and NSF and viceversa, facilitating collaborative research and exchange.
International collaboration should foster interactions

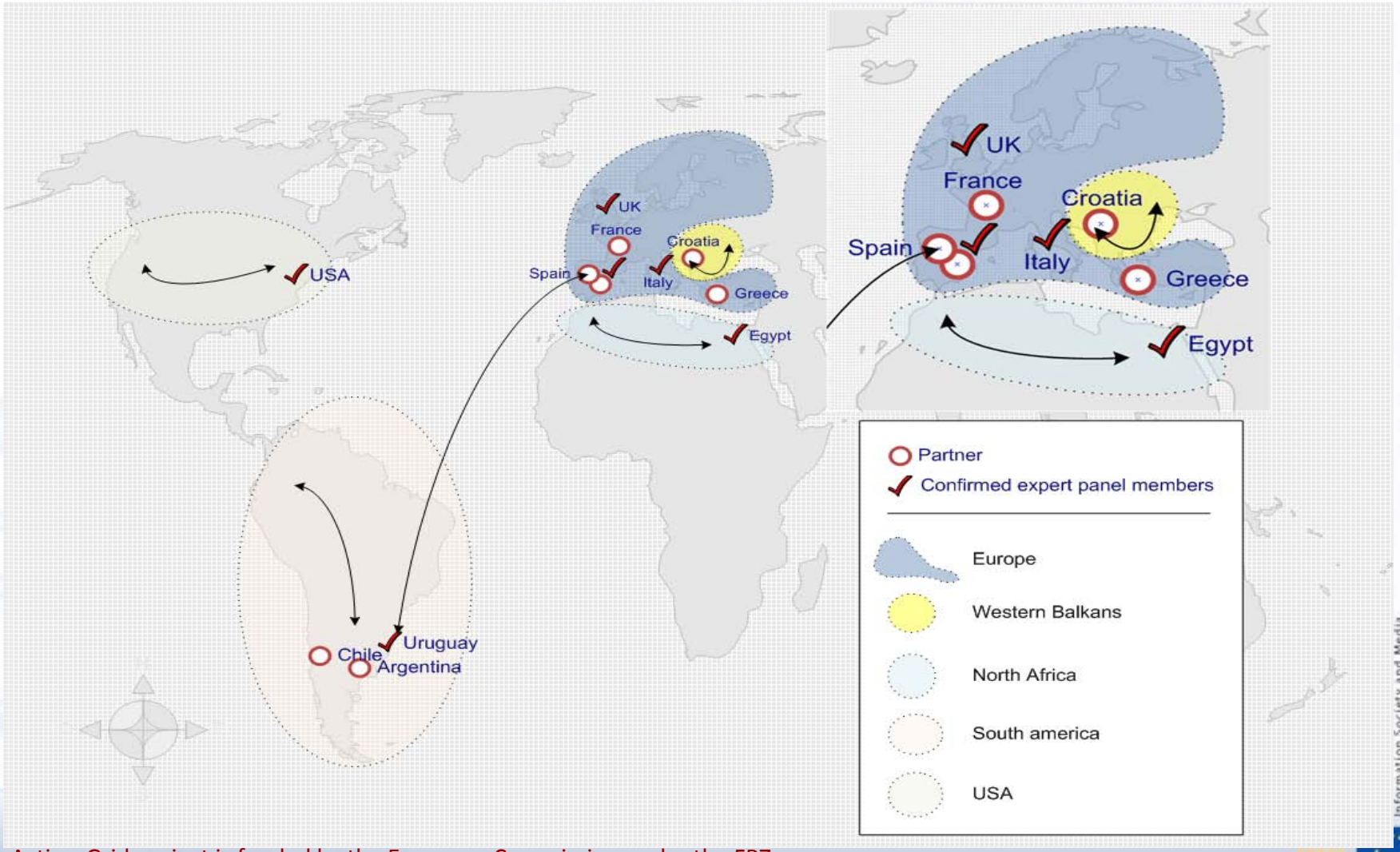
Nanotechnology and nanomedicine have important associated issues, like nanotoxicity —both animal and environmental—, and also ethical, cultural, media, entertainment, philosophical and even religious aspects

The EC has dealt with nano-related issues in directorates covering Nanotechnology and Health issues, but not in areas like ICT for Health. Opportunities for new programmes might arise, linking directorates and people working in the involved areas.

There are challenges for linking BMI and Nanoinformatics objectives, methods and tools that can be shared, exchanged and reused. Examples include databases, informatics tools, ontologies (biomedical and nano), services, etc.

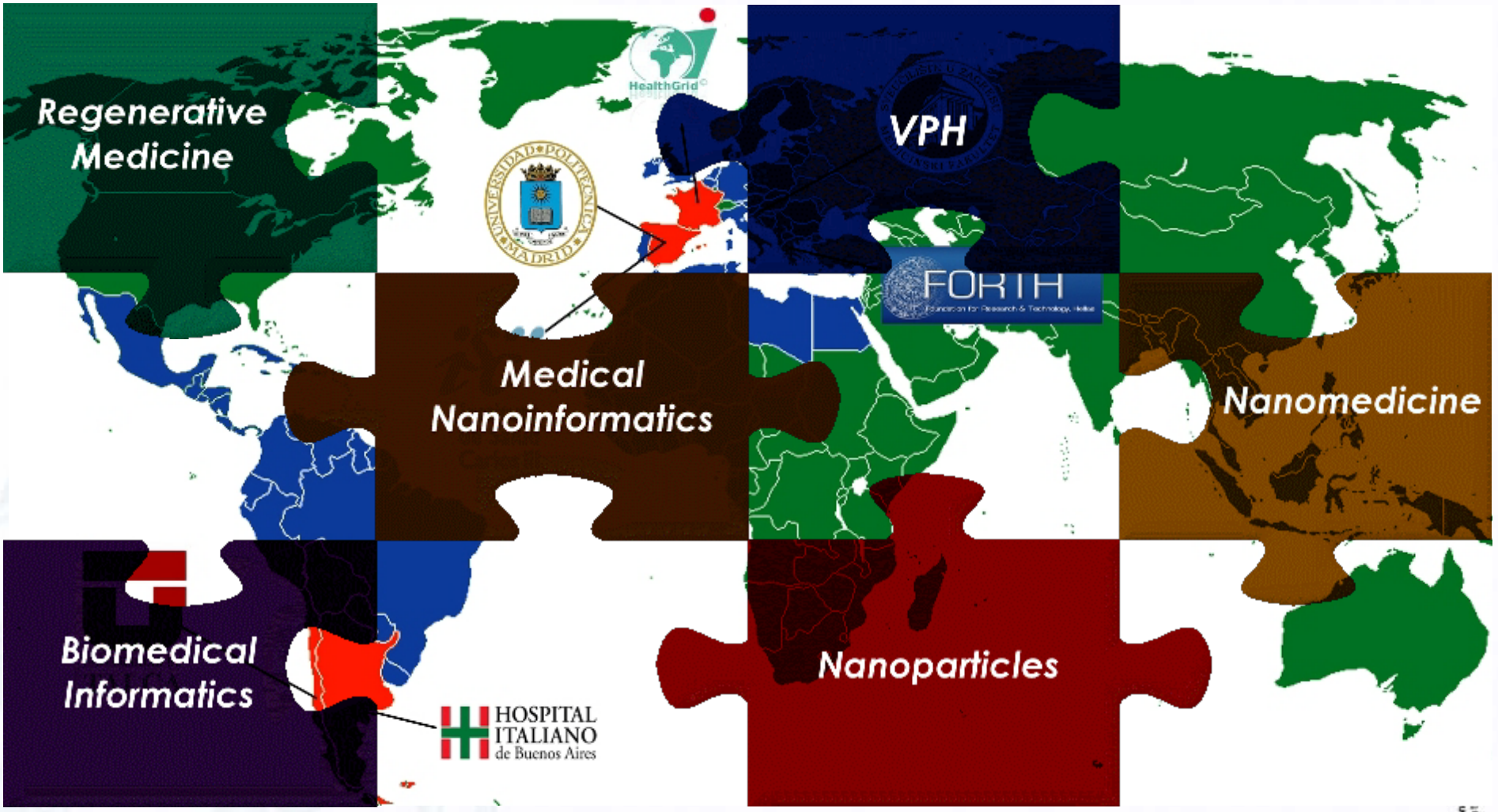
The Action-Grid project is funded by the European Commission under the 7th framework project

World-Wide Scope of ACTION-Grid: challenges for future international collaborations



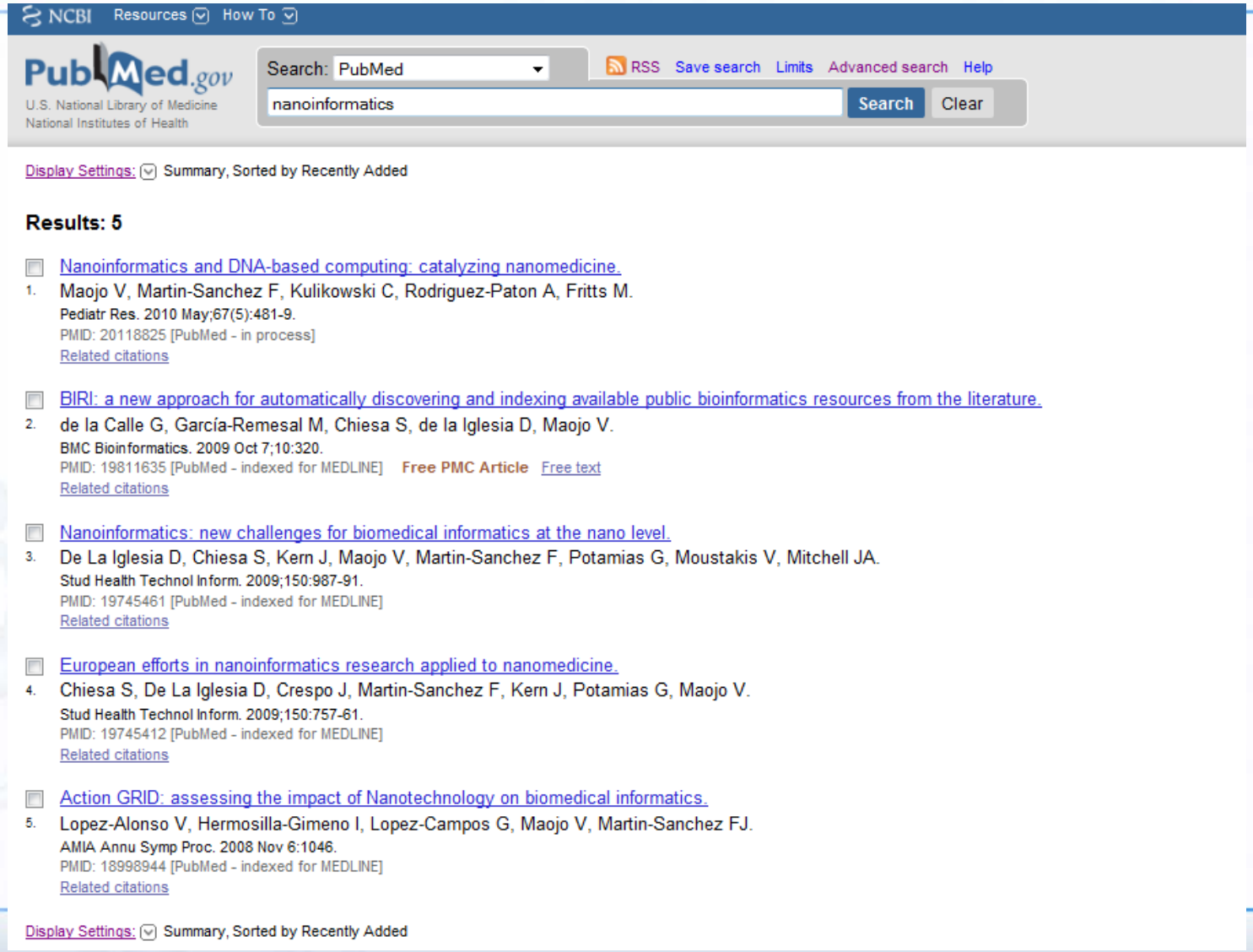
The Action-Grid project is funded by the European Commission under the FP7

Action-Grid: a finally (hopefully) coherent global, scientific vision



The Action-Grid project is funded by the European Commission under the FP7

Medline: 5 (+1) papers retrieved with the term “Nanoinformatics”



NCBI Resources How To

PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed | RSS | Save search | Limits | Advanced search | Help

nanoinformatics | Search | Clear

Display Settings: Summary, Sorted by Recently Added

Results: 5

1. Nanoinformatics and DNA-based computing: catalyzing nanomedicine.
Maojo V, Martin-Sanchez F, Kulikowski C, Rodriguez-Paton A, Fritts M.
Pediatr Res. 2010 May;67(5):481-9.
PMID: 20118825 [PubMed - in process]
Related citations

2. BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature.
de la Calle G, García-Remesal M, Chiesa S, de la Iglesia D, Maojo V.
BMC Bioinformatics. 2009 Oct 7;10:320.
PMID: 19811635 [PubMed - indexed for MEDLINE] Free PMC Article Free text
Related citations

3. Nanoinformatics: new challenges for biomedical informatics at the nano level.
De La Iglesia D, Chiesa S, Kern J, Maojo V, Martin-Sanchez F, Potamias G, Moustakis V, Mitchell JA.
Stud Health Technol Inform. 2009;150:987-91.
PMID: 19745461 [PubMed - indexed for MEDLINE]
Related citations

4. European efforts in nanoinformatics research applied to nanomedicine.
Chiesa S, De La Iglesia D, Crespo J, Martin-Sanchez F, Kern J, Potamias G, Maojo V.
Stud Health Technol Inform. 2009;150:757-61.
PMID: 19745412 [PubMed - indexed for MEDLINE]
Related citations

5. Action GRID: assessing the impact of Nanotechnology on biomedical informatics.
Lopez-Alonso V, Hermosilla-Gimeno I, Lopez-Campos G, Maojo V, Martin-Sanchez FJ.
AMIA Annu Symp Proc. 2008 Nov 6:1046.
PMID: 18998944 [PubMed - indexed for MEDLINE]
Related citations

Display Settings: Summary, Sorted by Recently Added

The Action-Grid project is funded by the European Commission under the FP7

Pediatric Research

Wolters Kluwer Health | Lippincott Williams & Wilkins

Pediatric Research:
May 2010 - Volume 67 - Issue 5 - pp 481-489
doi: 10.1203/PDR.0b013e3181d6245e
Nanopediatrics Review Articles: Improved Computer Technology

Nanoinformatics and DNA-Based Computing: Catalyzing Nanomedicine

MAOJO, VICTOR; MARTIN-SANCHEZ, FERNANDO; KULIKOWSKI, CASIMIR; RODRIGUEZ-PATON, ALFONSO; FRITTS, MARTIN

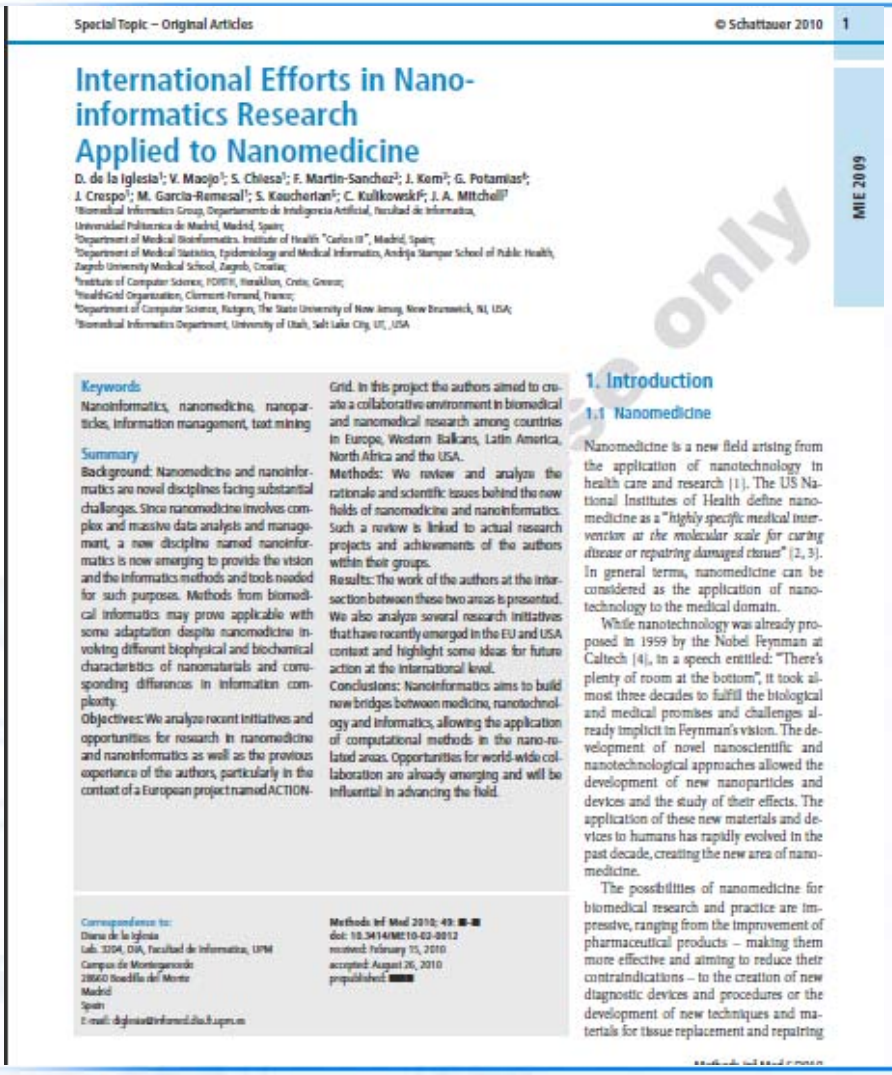
FREE

Article Outline

Author Information

Departamento de Inteligencia Artificial [V.M.], Universidad Politecnica de Madrid, Madrid 28660 Spain; Medical Bioinformatics Department [F.M.-S.], National Institute of Health "Carlos III," Madrid 28220, Spain; Department of Computer Science [C.K.], Rutgers, State University of New Jersey

And the sixth one is coming…



Special Topic – Original Articles © Schattauer 2010 1

MIE 2009

International Efforts in Nano-informatics Research Applied to Nanomedicine

D. de la Iglesia[1]; V. Maojo[1]; S. Chiesa[1]; F. Martin-Sanchez[2]; J. Kern[3]; G. Potamias[4]; J. Crespo[5]; M. Garcia-Remesal[1]; S. Kouchetian[6]; C. Kulikowski[6]; J. A. Mitchell[7]
[1]Biomedical Informatics Group, Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain;
[2]Department of Medical Bioinformatics, Institute of Health "Carlos III", Madrid, Spain;
[3]Department of Medical Statistics, Epidemiology and Medical Informatics, Andrija Stampar School of Public Health, Zagreb University Medical School, Zagreb, Croatia;
[4]Institute of Computer Science, FORTH, Heraklion, Crete, Greece;
[5]HealthGrid Organization, Clermont-Ferrand, France;
[6]Department of Computer Science, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA;
[7]Biomedical Informatics Department, University of Utah, Salt Lake City, UT, USA

Keywords
Nanoinformatics, nanomedicine, nanoparticles, information management, text mining

Summary
Background: Nanomedicine and nanoinformatics are novel disciplines facing substantial challenges. Since nanomedicine involves complex and massive data analysis and management, a new discipline named nanoinformatics is now emerging to provide the vision and the informatics methods and tools needed for such purposes. Methods from biomedical informatics may prove applicable with some adaptation despite nanomedicine involving different biophysical and biochemical characteristics of nanomaterials and corresponding differences in information complexity.
Objectives: We analyze recent initiatives and opportunities for research in nanomedicine and nanoinformatics as well as the previous experience of the authors, particularly in the context of a European project named ACTION-Grid. In this project the authors aimed to create a collaborative environment in biomedical and nanomedical research among countries in Europe, Western Balkans, Latin America, North Africa and the USA.
Methods: We review and analyze the rationale and scientific issues behind the new fields of nanomedicine and nanoinformatics. Such a review is linked to actual research projects and achievements of the authors within their groups.
Results: The work of the authors at the intersection between these two areas is presented. We also analyze several research initiatives that have recently emerged in the EU and USA context and highlight some ideas for future action at the international level.
Conclusions: Nanoinformatics aims to build new bridges between medicine, nanotechnology and informatics, allowing the application of computational methods in the nano-related areas. Opportunities for world-wide collaboration are already emerging and will be influential in advancing the field.

Correspondence to:
Diana de la Iglesia
Lab. 3204, DIA, Facultad de Informática, UPM
Campus de Montegancedo
28660 Boadilla del Monte
Madrid
Spain
E-mail: diglesia@fmformed.dia.fi.upm.es

Methods Inf Med 2010; 49: ■-■
doi: 10.3414/ME10-02-0012
received: February 15, 2010
accepted: August 26, 2010
prepublished: ■■■

1. Introduction

1.1 Nanomedicine

Nanomedicine is a new field arising from the application of nanotechnology in health care and research [1]. The US National Institutes of Health define nanomedicine as a "*highly specific medical intervention at the molecular scale for curing disease or repairing damaged tissues*" [2, 3]. In general terms, nanomedicine can be considered as the application of nanotechnology to the medical domain.

While nanotechnology was already proposed in 1959 by the Nobel Feynman at Caltech [4], in a speech entitled: "There's plenty of room at the bottom", it took almost three decades to fulfill the biological and medical promises and challenges already implicit in Feynman's vision. The development of novel nanoscientific and nanotechnological approaches allowed the development of new nanoparticles and devices and the study of their effects. The application of these new materials and devices to humans has rapidly evolved in the past decade, creating the new area of nanomedicine.

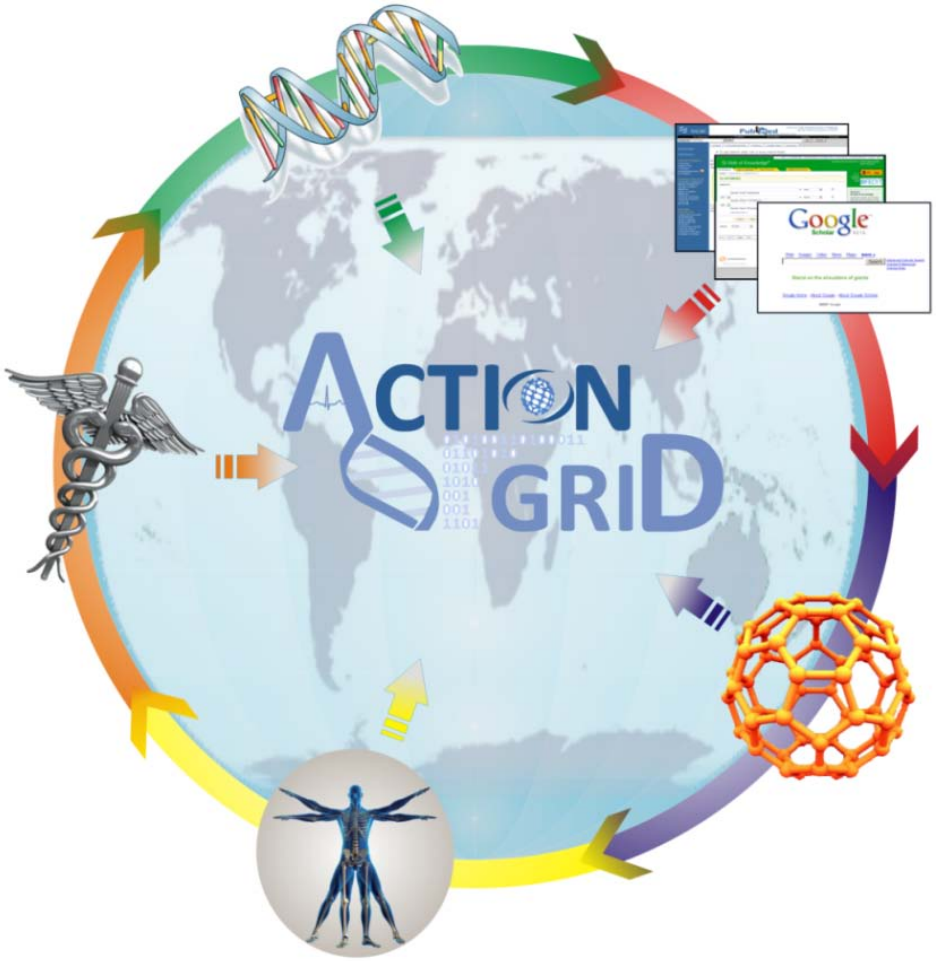
The possibilities of nanomedicine for biomedical research and practice are impressive, ranging from the improvement of pharmaceutical products – making them more effective and aiming to reduce their contraindications – to the creation of new diagnostic devices and procedures or the development of new techniques and materials for tissue replacement and repairing

In press in *Methods of Information in Medicine*

The Action-Grid project is funded by the European Commission under the 7th framework project

Victor Maojo

Research topics at the UPM group



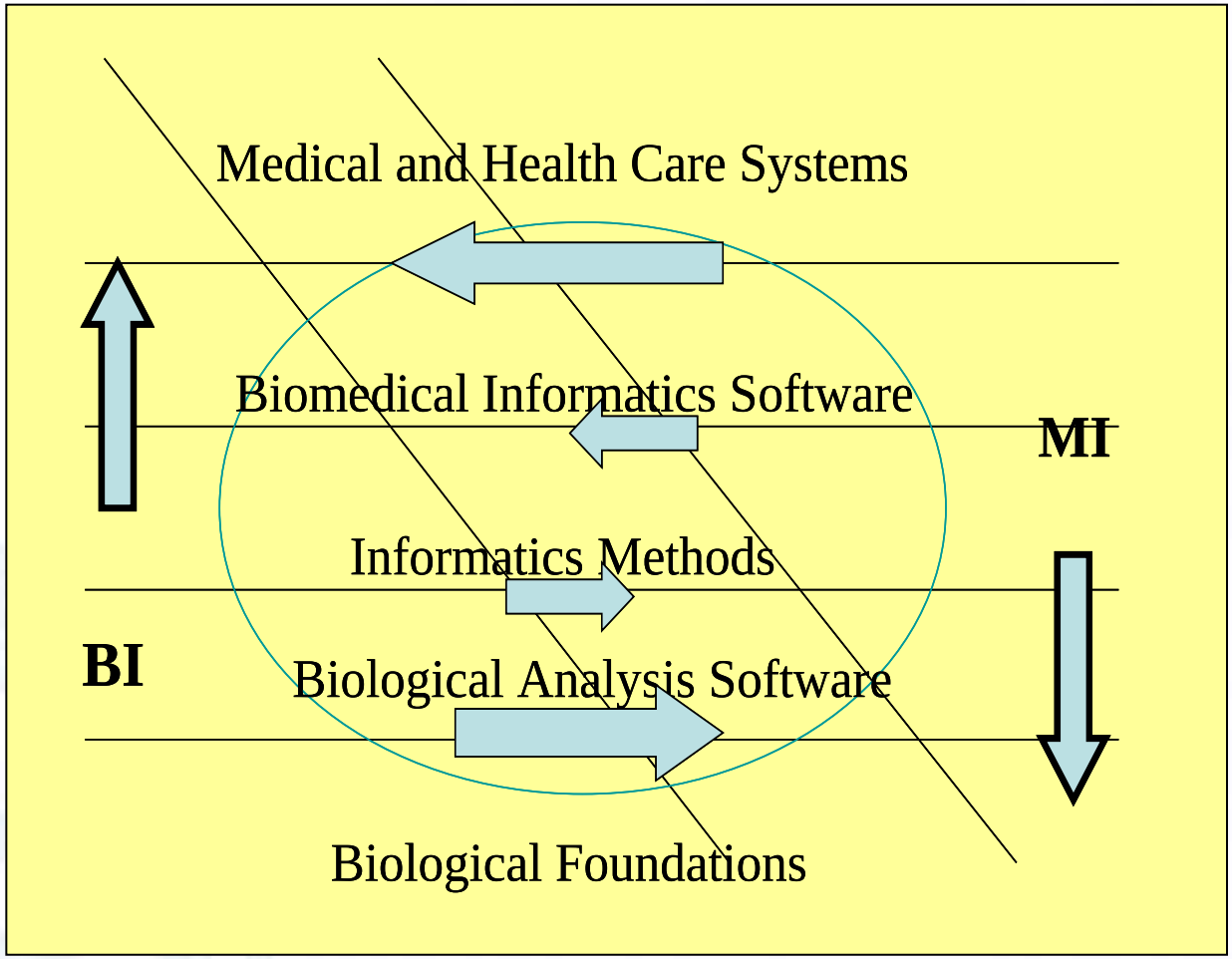
Nanoinformatics: a challenging research context, where former experiences can be applied

Novel idea (Nanoinformatics), a (possible) new discipline linked to Biomedical Informatics and Nanomedicine, with almost no previous publications at all

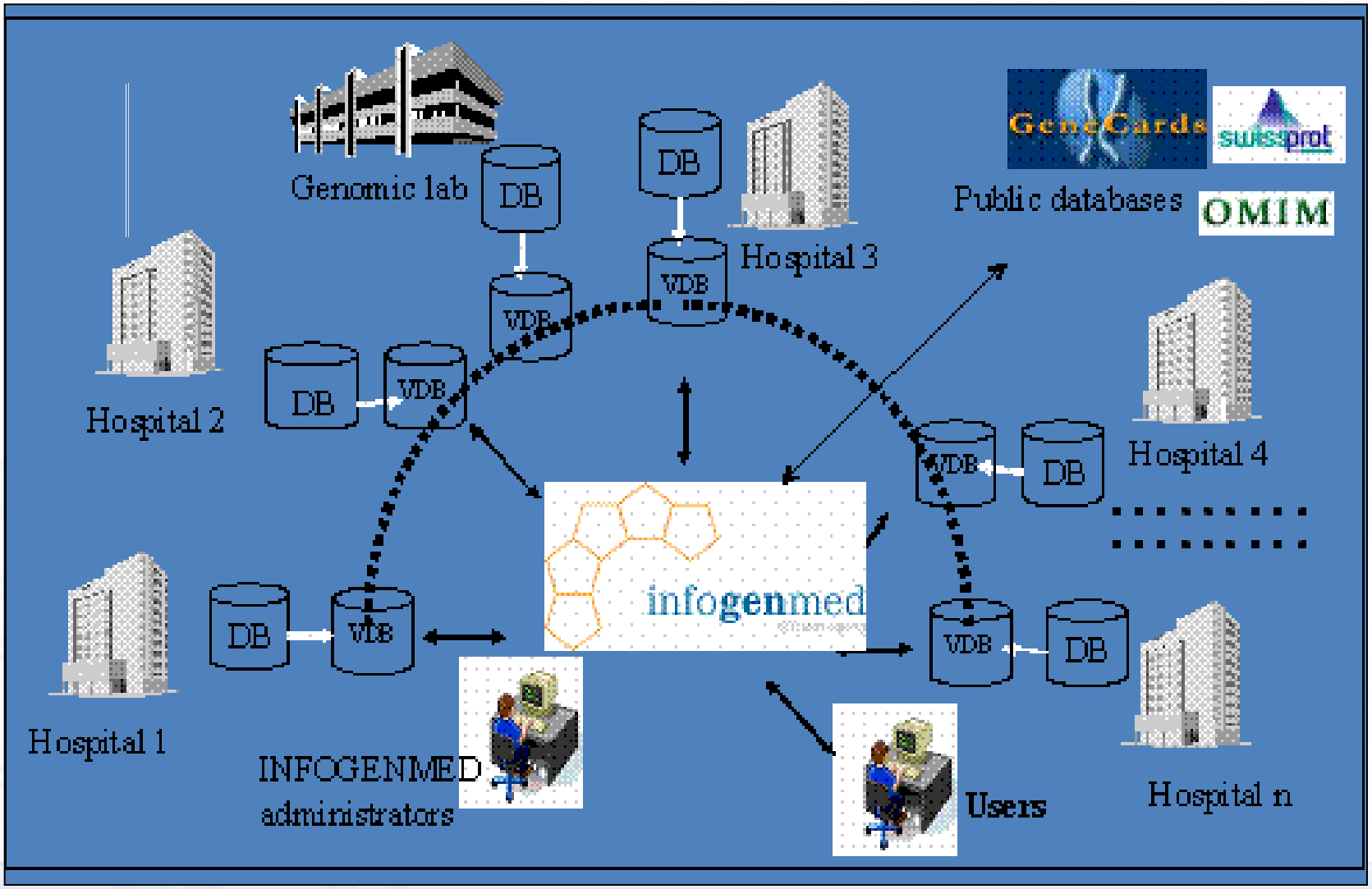
Amazing number of related problems to Biomedical Informatics, from an informatics perspective, with deep differences but also similarities

The Action-Grid project is funded by the European Commission under the FP7

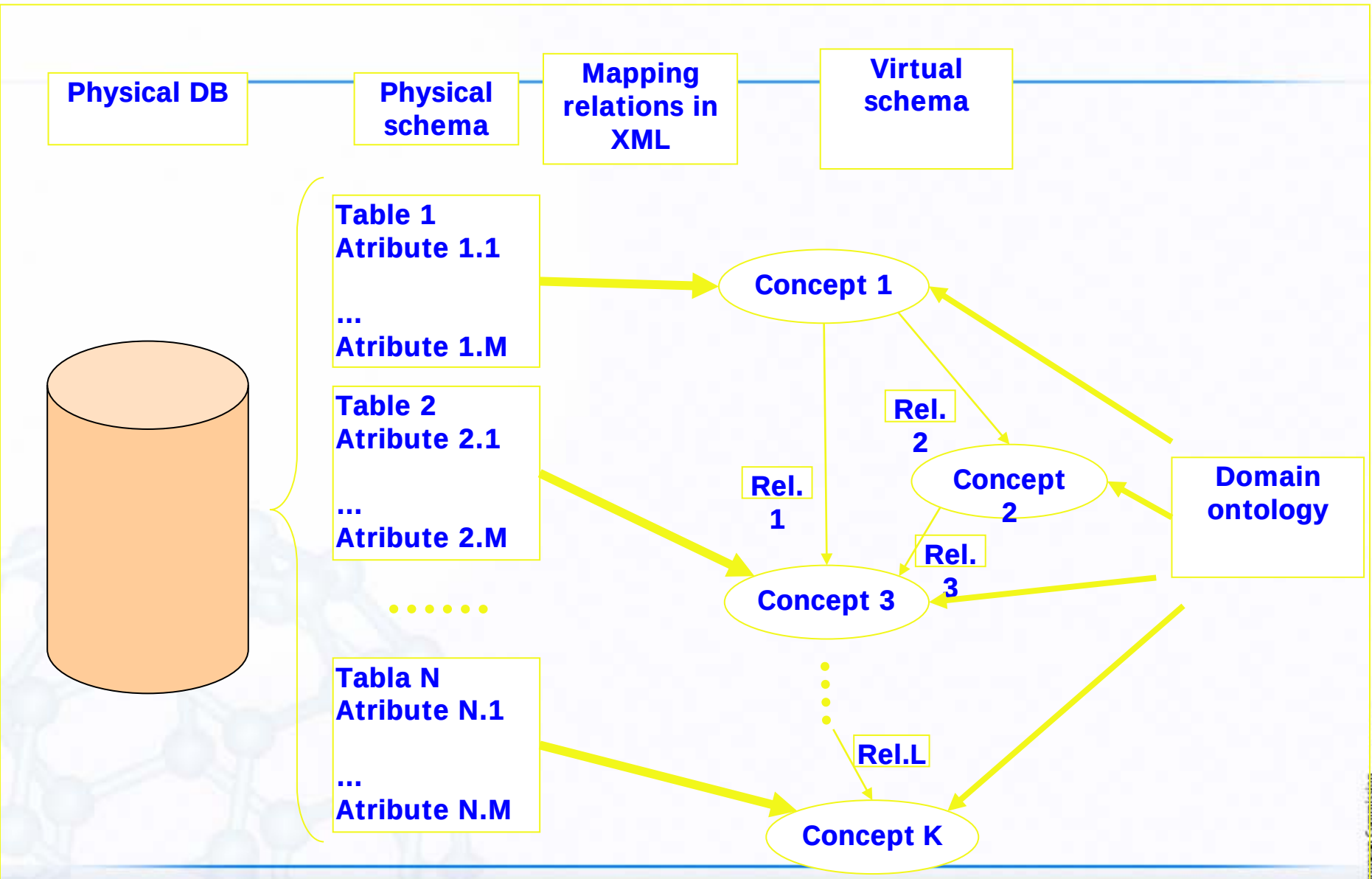
Challenges for Medical Informatics /Bioinformatics interactions (Maojo and Kulikowski, JAMIA, Nov. 2003)



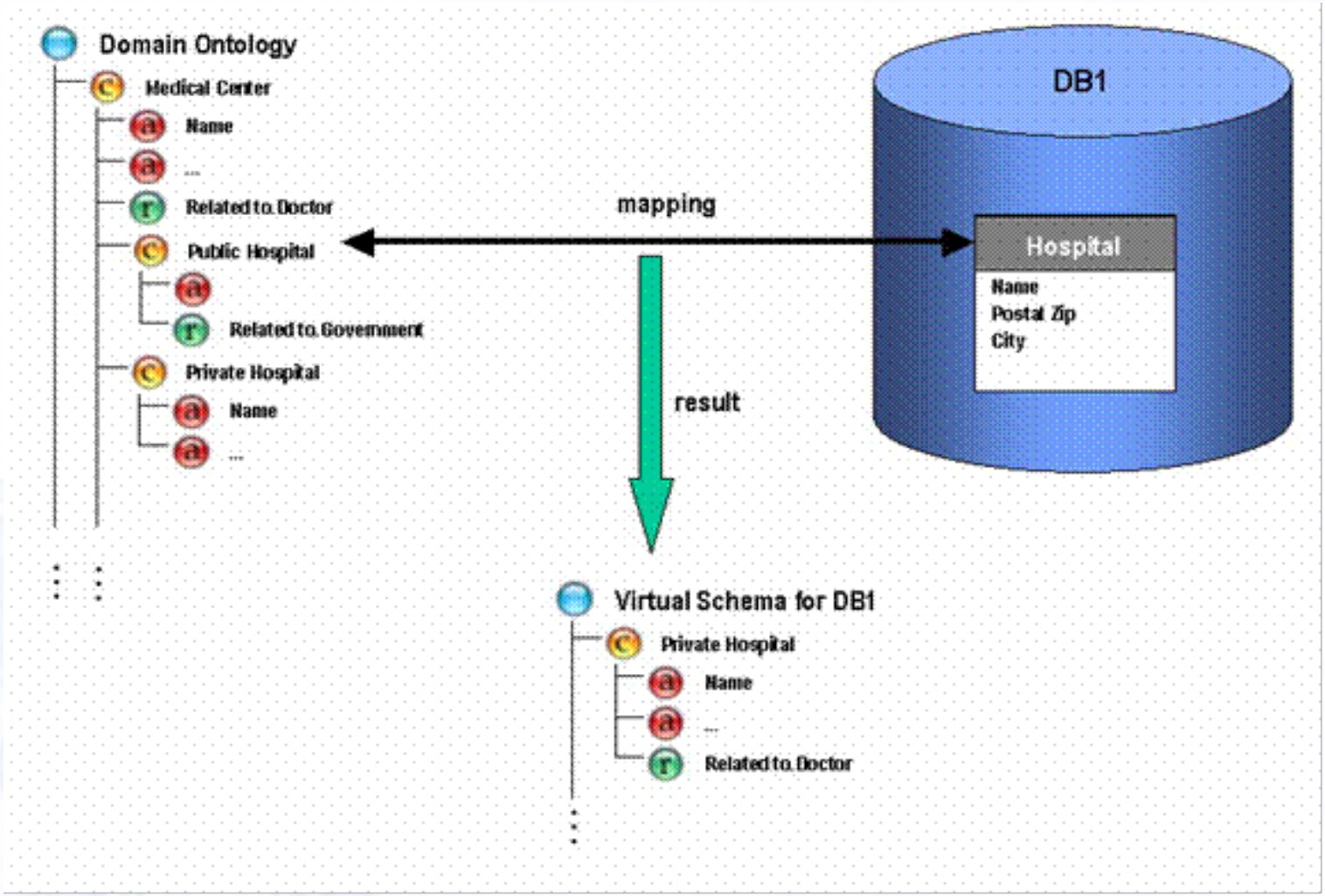
Our approach to clinical and -omics (heterogeneous) database integration



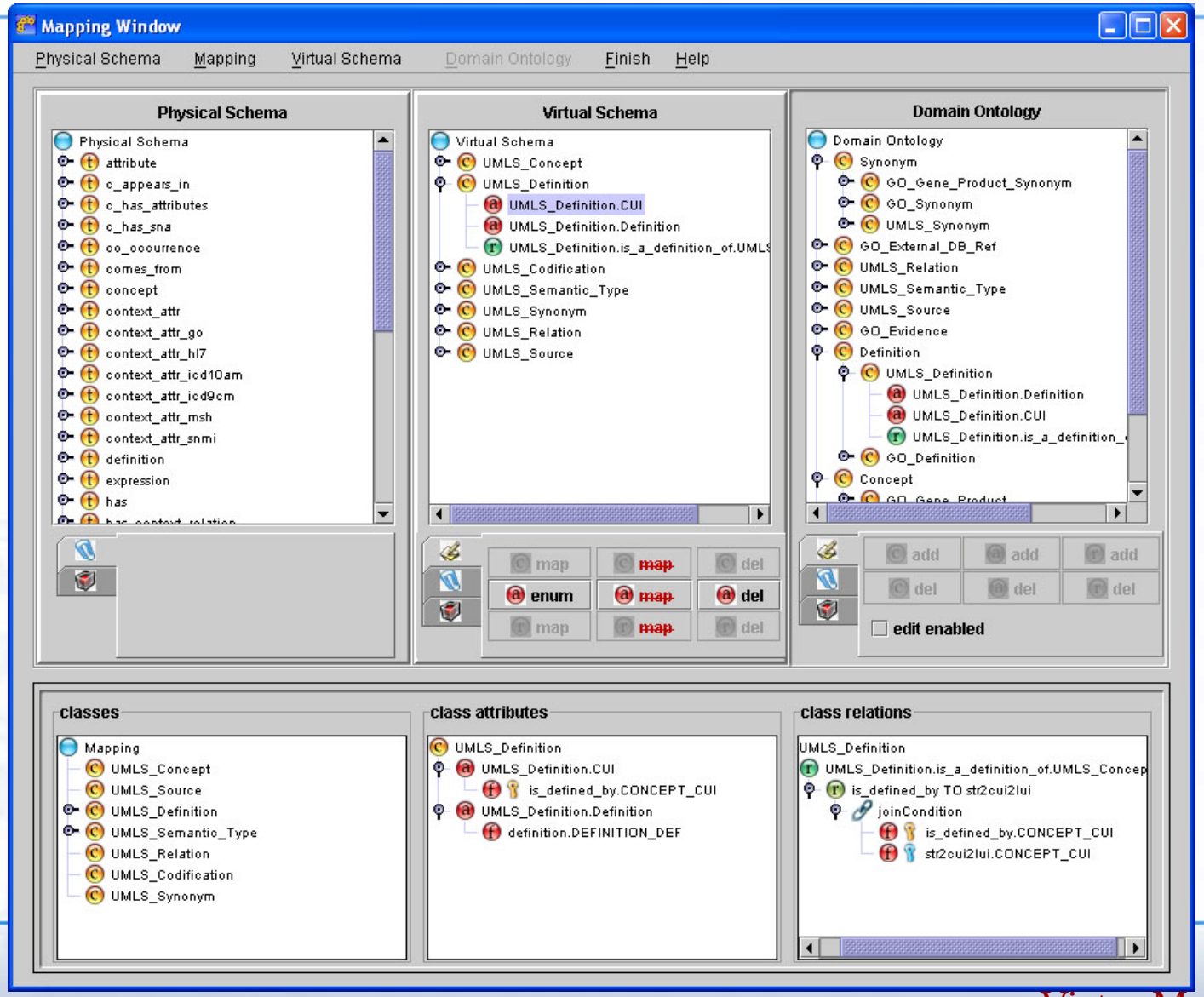
Mapping schema



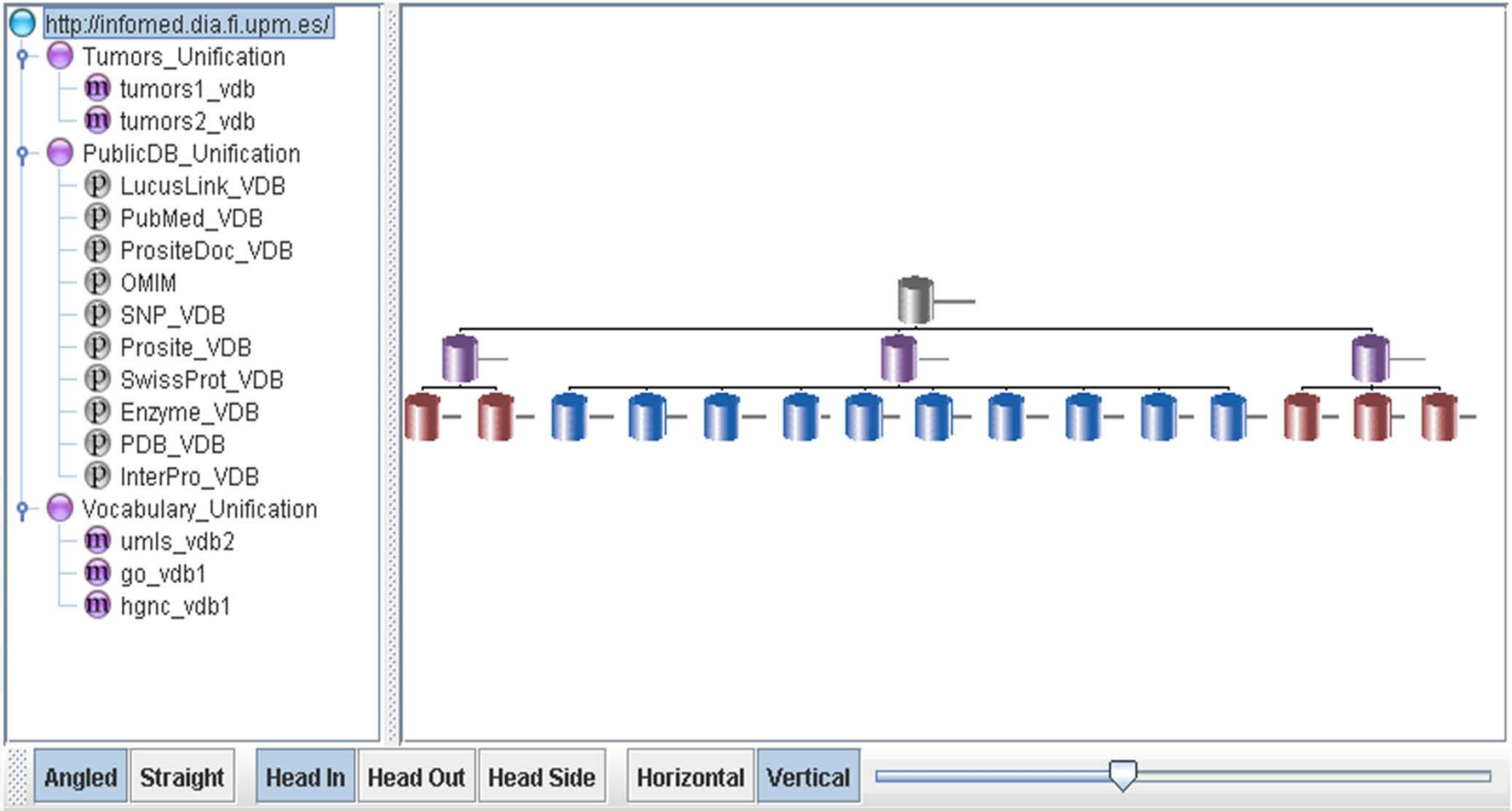
Homogeneization model



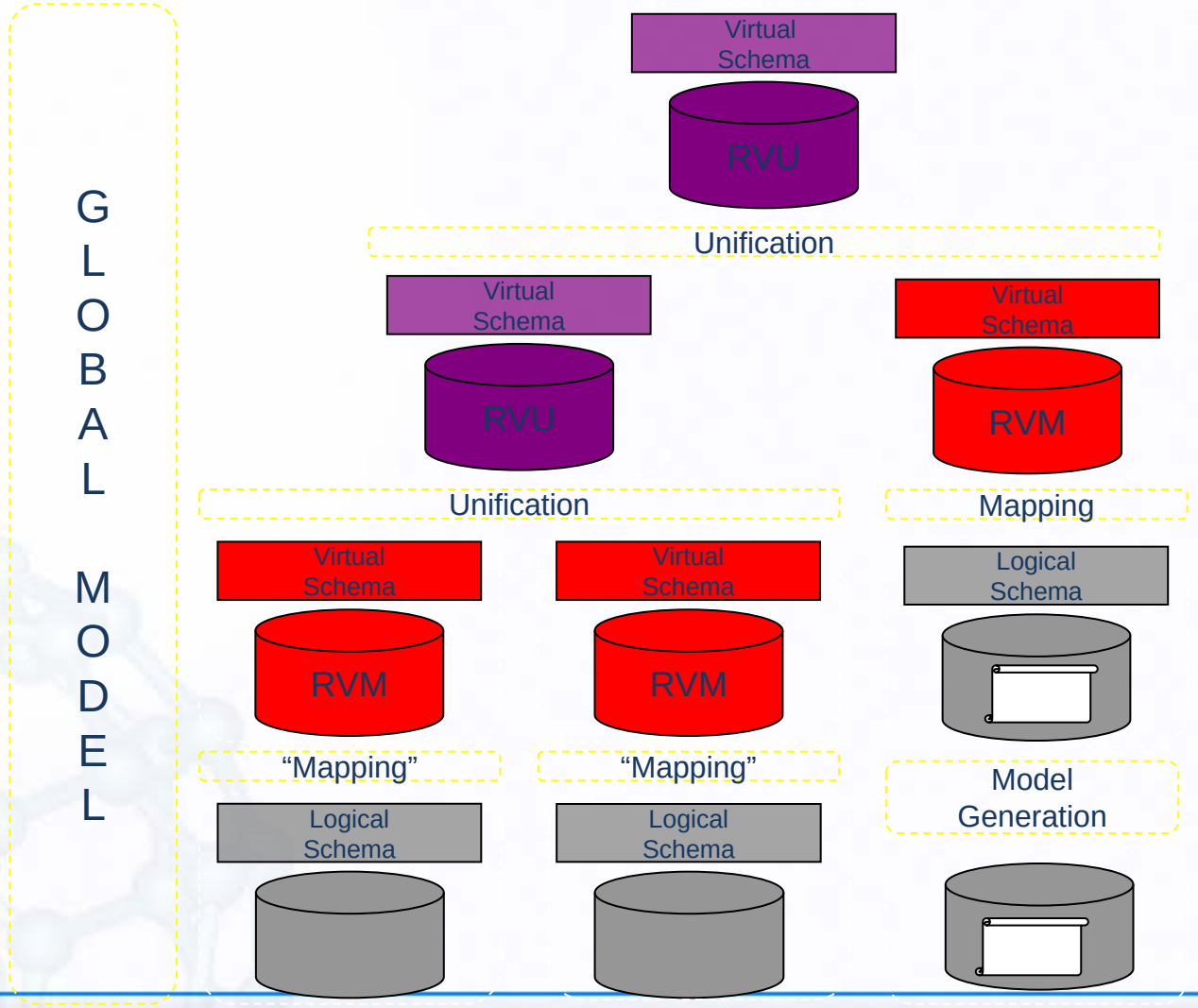
Mapping tools



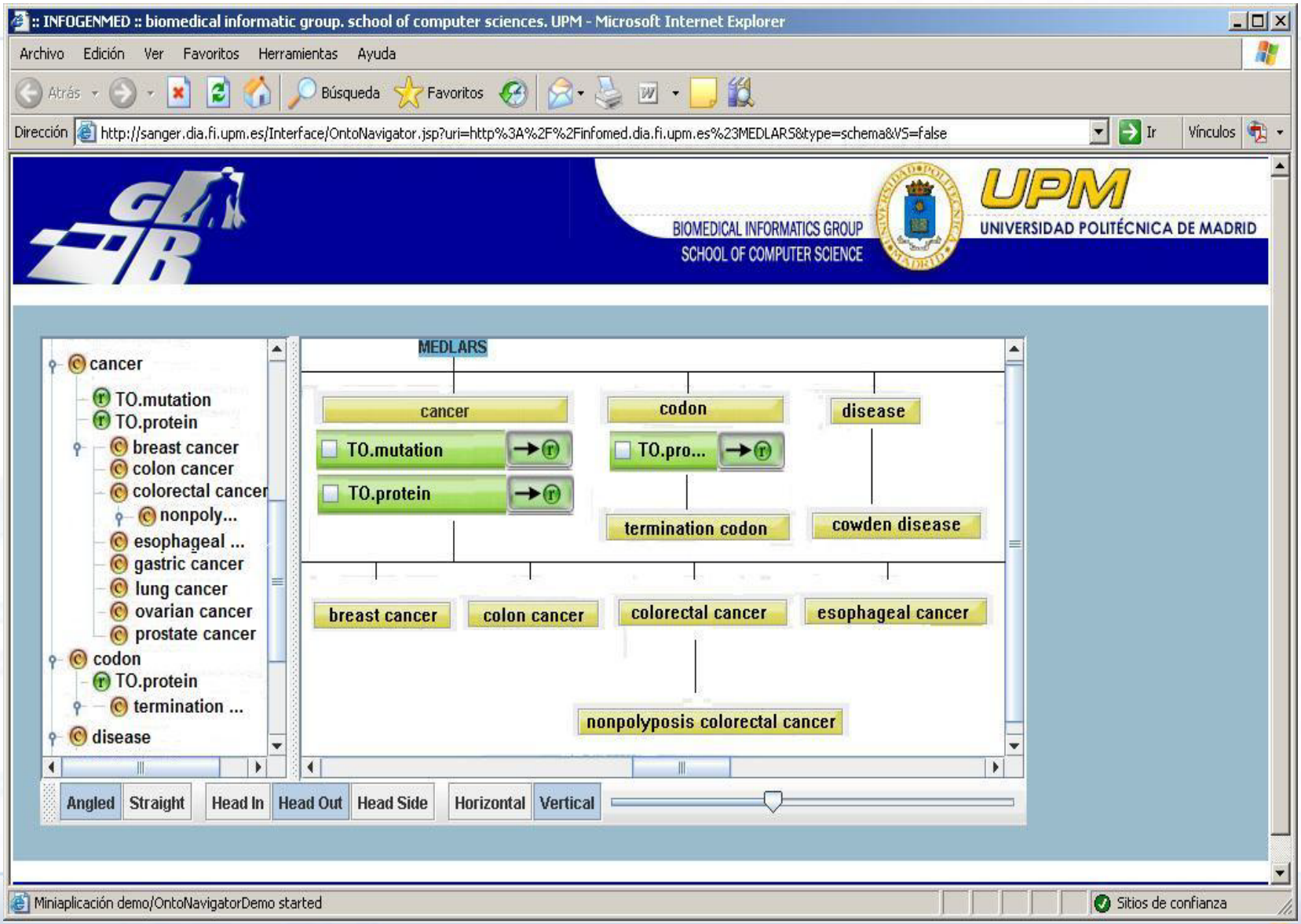
Clinical and Genomic database integration



Expanding to integration of structured and non-structured sources (ONTOFUSION)



Results

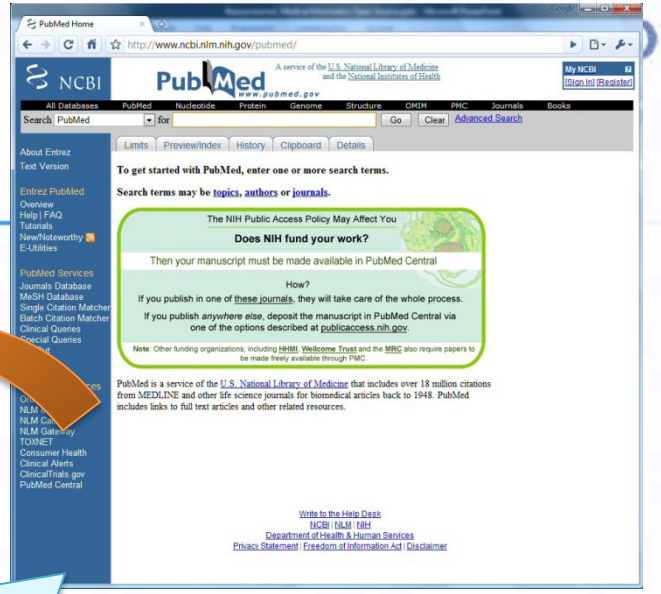

:: INFOGENMED :: biomedical informatic group. school of computer sciences. UPM - Microsoft Internet Explorer
Archivo Edición Ver Favoritos Herramientas Ayuda
Atrás Búsqueda Favoritos
Dirección http://sanger.dia.fi.upm.es/Interface/OntoNavigator.jsp?uri=http%3A%2F%2Finfomed.dia.fi.upm.es%23MEDLARS&type=schema&WS=false
Ir Vínculos
BIOMEDICAL INFORMATICS GROUP
SCHOOL OF COMPUTER SCIENCE
UPM
UNIVERSIDAD POLITÉCNICA DE MADRID
cancer
TO.mutation
TO.protein
breast cancer
colon cancer
colorectal cancer
nonpoly...
esophageal ...
gastric cancer
lung cancer
ovarian cancer
prostate cancer
codon
TO.protein
termination ...
disease
MEDLARS
cancer
codon
disease
TO.mutation
TO.pro...
TO.protein
termination codon
cowden disease
breast cancer
colon cancer
colorectal cancer
esophageal cancer
nonpolyposis colorectal cancer
Angled
Straight
Head In
Head Out
Head Side
Horizontal
Vertical
Miniaplicación demo/OntoNavigatorDemo started
Sitios de confianza

Applications: integration of structured and non-structured sources (ONTOFUSION)

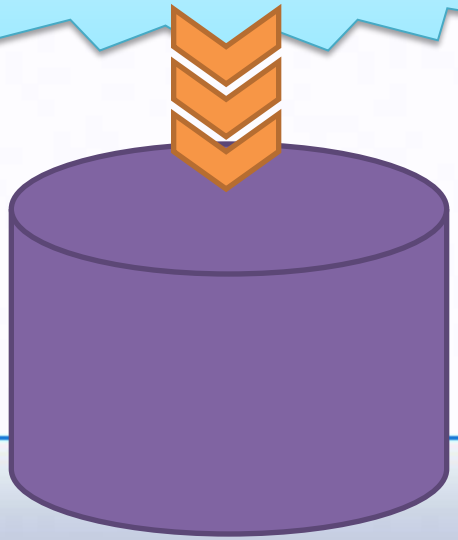
- Unified Model
  - 257 concepts.
  - 106 hierarchical relationships
  - 425 “ad-hoc” relationships

Subset of the Unified Model

Victor Maojo



Automatic Extraction of Resources' Information from the literature (like PubMed & PMC)

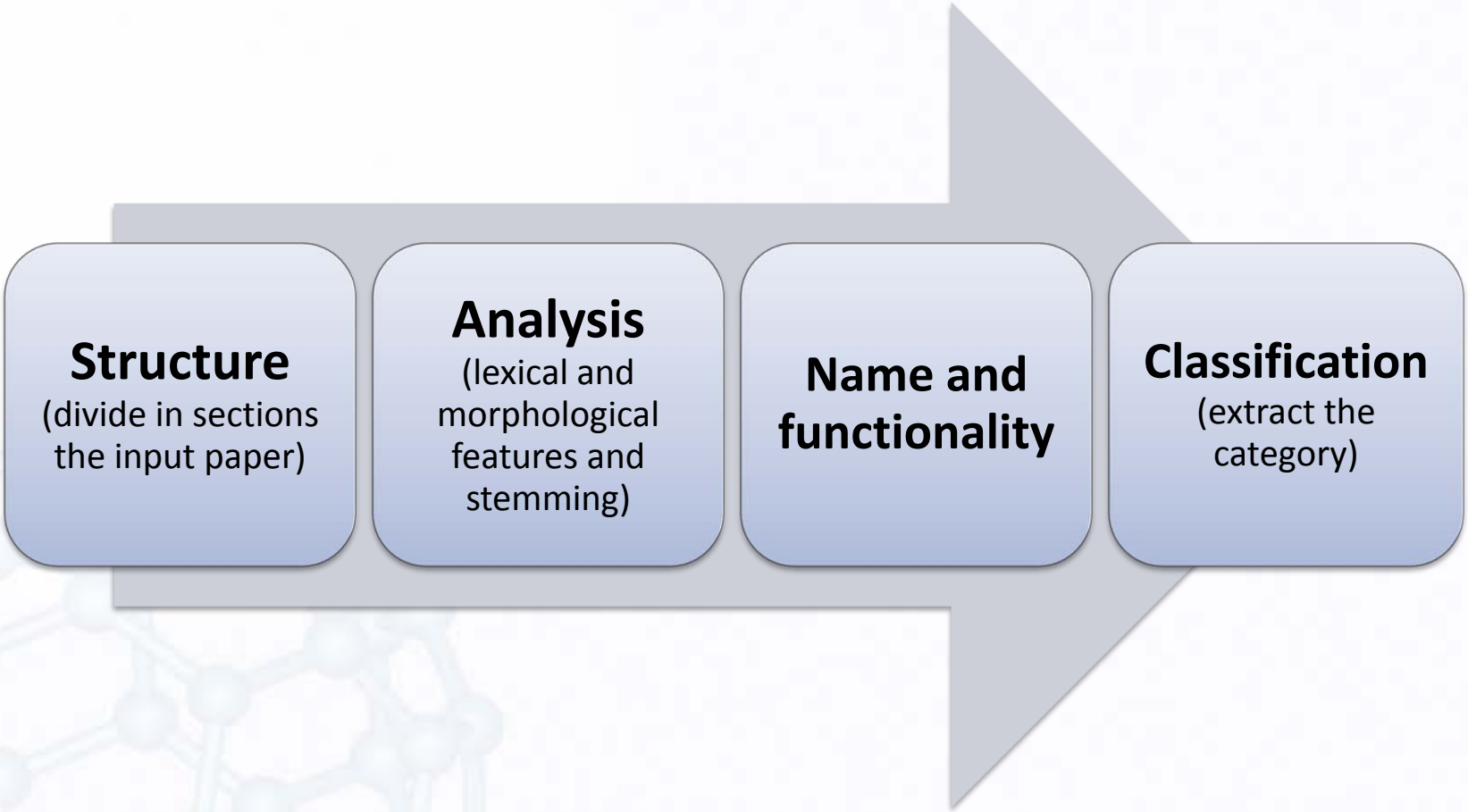


Public Open Source
BioMedical
Informatics
Resources Inventory

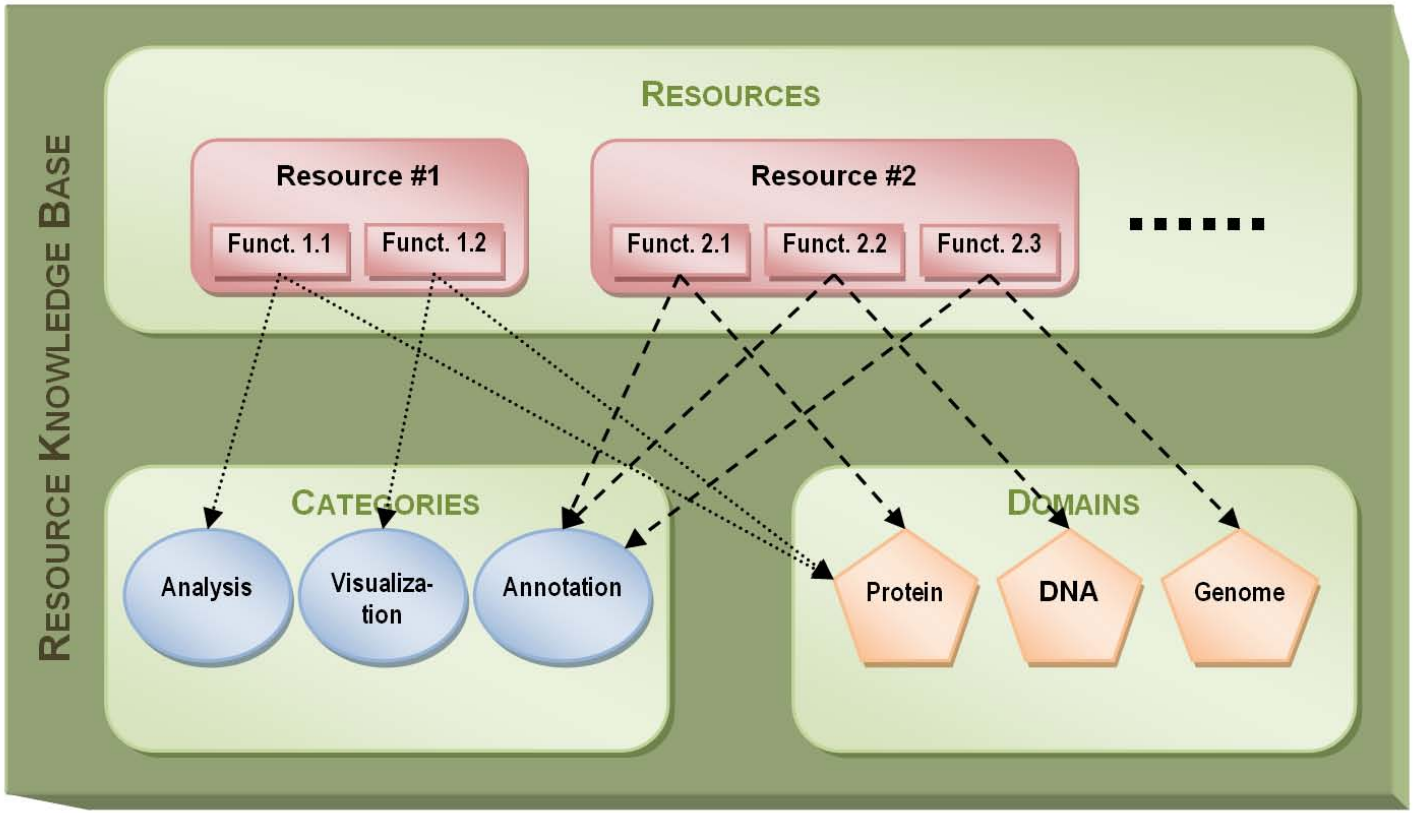
European Commission
Information Society and Media

Victor Maojo

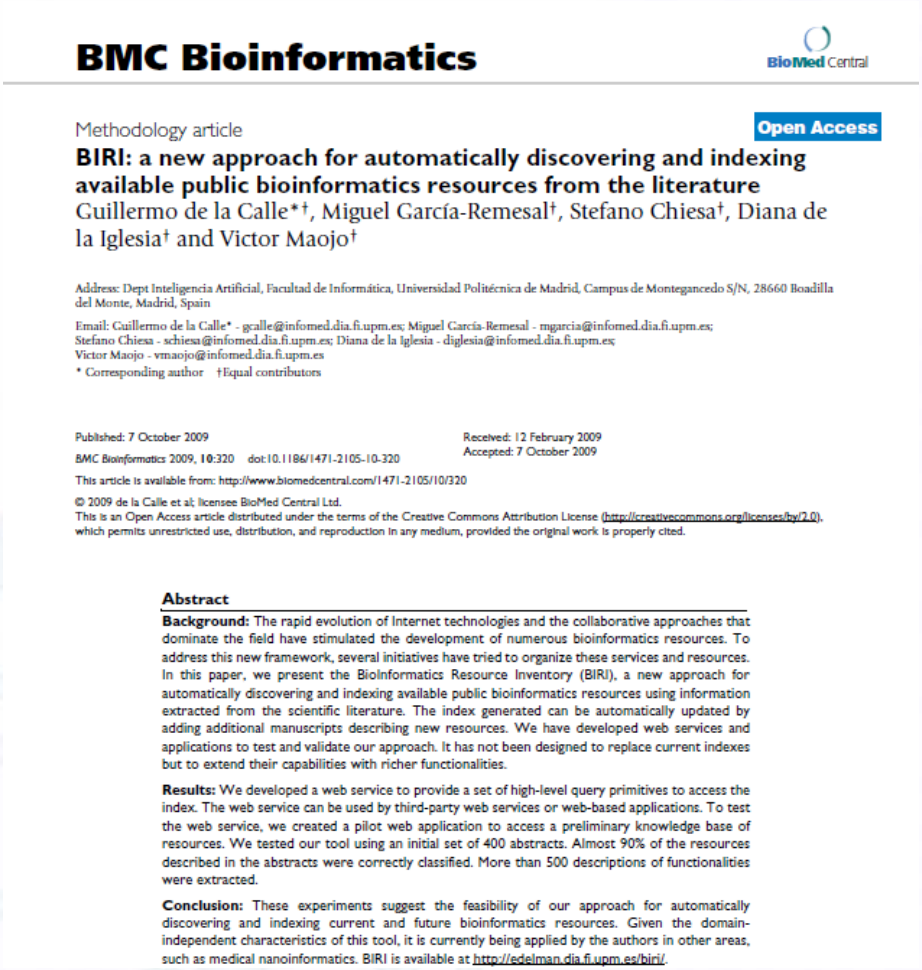
The Analysis process



How to build the Resource Knowledge Base



Towards a bioinformatics resourceome



BMC Bioinformatics

BioMed Central

Methodology article

Open Access

BIRI: a new approach for automatically discovering and indexing available public bioinformatics resources from the literature

Guillermo de la Calle*†, Miguel García-Remesal†, Stefano Chiesa†, Diana de la Iglesia† and Victor Maojo†

Address: Dept Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo S/N, 28660 Boadilla del Monte, Madrid, Spain

Email: Guillermo de la Calle* - gcalle@infomed.dia.fi.upm.es; Miguel García-Remesal - mgarcia@infomed.dia.fi.upm.es; Stefano Chiesa - schiesa@infomed.dia.fi.upm.es; Diana de la Iglesia - diglesia@infomed.dia.fi.upm.es; Victor Maojo - vmaojo@infomed.dia.fi.upm.es

* Corresponding author †Equal contributors

Published: 7 October 2009

BMC Bioinformatics 2009, **10**:320 doi:10.1186/1471-2105-10-320

Received: 12 February 2009
Accepted: 7 October 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/320

Abstract

Background: The rapid evolution of Internet technologies and the collaborative approaches that dominate the field have stimulated the development of numerous bioinformatics resources. To address this new framework, several initiatives have tried to organize these services and resources. In this paper, we present the BioInformatics Resource Inventory (BIRI), a new approach for automatically discovering and indexing available public bioinformatics resources using information extracted from the scientific literature. The index generated can be automatically updated by adding additional manuscripts describing new resources. We have developed web services and applications to test and validate our approach. It has not been designed to replace current indexes but to extend their capabilities with richer functionalities.

Results: We developed a web service to provide a set of high-level query primitives to access the index. The web service can be used by third-party web services or web-based applications. To test the web service, we created a pilot web application to access a preliminary knowledge base of resources. We tested our tool using an initial set of 400 abstracts. Almost 90% of the resources described in the abstracts were correctly classified. More than 500 descriptions of functionalities were extracted.

Conclusion: These experiments suggest the feasibility of our approach for automatically discovering and indexing current and future bioinformatics resources. Given the domain-independent characteristics of this tool, it is currently being applied by the authors in other areas, such as medical nanoinformatics. BIRI is available at http://edelman.dia.fi.upm.es/biri/.

BIRI: A new method for the automatic discovery and indexing of bioinformatics references from the literature, designed to create a repository of resources

The Action-Grid project is funded by the European Commission under the FP7

BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 21 2010, pages 2801–2802
doi:10.1093/bioinformatics/btq520

Databases and ontologies

Advance Access publication September 9, 2010

PubDNA Finder: a web database linking full-text articles to sequences of nucleic acids

Miguel García-Remesal[1,2,*], Alejandro Cuevas[2], David Pérez-Rey[1,2], Luis Martín[2], Alberto Anguita[2], Diana de la Iglesia[2], Guillermo de la Calle[2], José Crespo[2,3] and Víctor Maojo[1,2]

[1]Departamento de Inteligencia Artificial, Facultad de Informática, [2]Biomedical Informatics Group, Facultad de Informática and [3]Departamento de Lenguajes, Sistemas Informáticos e Ingeniería del Software, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegacedo S/N, 28660 Boadilla del Monte, Madrid, Spain

Associate Editor: John Quackenbush

ABSTRACT

Summary: PubDNA Finder is an online repository that we have created to link PubMed Central manuscripts to the sequences of nucleic acids appearing in them. It extends the search capabilities provided by PubMed Central by enabling researchers to perform advanced searches involving sequences of nucleic acids. This includes, among other features (i) searching for papers mentioning one or more specific sequences of nucleic acids and (ii) retrieving the genetic sequences appearing in different articles. These additional query capabilities are provided by a searchable index that we created by using the full text of the 176 672 papers available at PubMed Central at the time of writing and the sequences of nucleic acids appearing in them. To automatically extract the genetic sequences occurring in each paper, we used an original method we have developed. The database is updated monthly by automatically connecting to the PubMed Central FTP site to retrieve and index new manuscripts. Users can query the database via the web interface provided.

Availability: PubDNA Finder can be freely accessed at http://servet.dia.fi.upm.es:8080/pubdnafinder

Contact: mgarcia@infomed.dia.fi.upm.es

Received on July 16, 2010; revised on August 27, 2010; accepted on September 4, 2010

1 INTRODUCTION

The biological literature is the main source of information reporting empirically validated genetic sequences, such as for instance PCR primers and probes. As result, researchers usually need to review the available literature to search for sequence data, which can be a hard and time-consuming task. PubMed Central is currently the main source of open-access full-text papers reporting genetic sequence data. However, the search engine provided by PubMed Central does not support researchers to retrieve papers containing the genetic sequences specified by the user, and to automatically identify and extract the sequences of nucleic acids mentioned in the retrieved articles.

PubDNA Finder is an online repository linking PubMed Central manuscripts to the different genetic sequences appearing in them. It extends the search capabilities provided by PubMed Central by allowing researchers to (i) retrieve all articles containing the genetic sequences specified by the user—featuring both exact and approximate matching; (ii) retrieve all the sequences appearing in the manuscripts matching a keyword-based query; and (iii) retrieve all articles matching a keyword-based query and containing the sequences specified by the user. PubDNA Finder currently contains the 176 672 papers available from PubMed Central at the time of writing. The database is automatically updated on a monthly basis to retrieve and index new manuscripts.

2 METHODS

To create the index, we downloaded all the 176 672 XML-formatted manuscripts available from the PubMed Central FTP site[1] at the time of writing. We used Apache Lucene[2] 3.0.1 to index the different documents based on the full text of the manuscripts and the genetic sequences appearing in each manuscript. The latter were automatically identified and extracted—together with the context in which they appeared—using a method created by the authors and reported elsewhere (García-Remesal *et al.*, 2010). The adopted method resorts to a rule-based expert system to automatically identify and extract the sequences of nucleic acids. To enable users to interactively query the developed index, we created a web interface.

3 FEATURES

Users can perform three different types of queries using PubDNA Finder, as described below.

3.1 Sequence-based queries

Sequence-based queries (SBQs) are aimed at retrieving all manuscripts containing one or more genetic sequences specified by the user. There are two different types of SBQs: simple and advanced. Simple SBQs are composed of one or more complete sequences linked by a single logical operator, such as 'retrieve all manuscripts containing either the sequence TATGGAAMAGATC-GGCGG or the sequence ATTGGCGAAGTCGGTAGG'.To launch this query, we would type the target sequences—one per line—in the text-box labeled with 'Sequences' (Fig. 1) and then we would select the OR logical operator in the 'Operator' combo box.

*To whom correspondence should be addressed.

[1]ftp://ftp.ncbi.nlm.nih.gov/pub/pmc
[2]http://lucene.apache.org/

BMC Bioinformatics

METHODOLOGY ARTICLE Open Access

A method for automatically extracting infectious disease-related primers and probes from the literature

Miguel García-Remesal[1,2*†], Alejandro Cuevas[2†], Victoria López-Alonso[2†], Guillermo López-Campos[2†], Guillermo de la Calle[2†], Diana de la Iglesia[2†], David Pérez-Rey[1,2†], José Crespo[2,3†], Fernando Martín-Sánchez[2†], Víctor Maojo[1,2†]

Abstract

Background: Primer and probe sequences are the main components of nucleic acid-based detection systems. Biologists use primers and probes for different tasks, some related to the diagnosis and prescription of infectious diseases. The biological literature is the main information source for empirically validated primer and probe sequences. Therefore, it is becoming increasingly important for researchers to navigate this important information. In this paper, we present a four-phase method for extracting and annotating primer/probe sequences from the literature. These phases are: (i) convert each document into a tree of paper sections, (2) detect the candidate sequences using a set of finite state machine-based recognizers, (3) refine problem sequences using a rule-based expert system, and (4) annotate the extracted sequences with their related organism/gene information.

Results: We tested our approach using a test set composed of 297 manuscripts. The extracted sequences and their organism/gene annotations were manually evaluated by a panel of molecular biologists. The results of the evaluation show that our approach is suitable for automatically extracting DNA sequences, achieving precision/recall rates of 97.98% and 95.77%, respectively. In addition, 76.66% of the detected sequences were correctly annotated with their organism name. The system also provided correct gene-related information for 46.18% of the sequences assigned a correct organism name.

Conclusions: We believe that the proposed method can facilitate routine tasks for biomedical researchers using molecular methods to diagnose and prescribe different infectious diseases. In addition, the proposed method can be expanded to detect and extract other biological sequences from the literature. The extracted information can also be used to readily update available primer/probe databases or to create new databases from scratch.

Background

Molecular technologies are used in routine clinical practice to identify microorganisms, and evaluate the presence of virulence factors, antibiotic resistance determinants and host-microbe interactions [1]. For instance, numerous nucleic acid assays have been developed [2] using hybridization or DNA extension techniques that include a wide range of technologies, such as polymerase chain reaction (PCR) methods [3], gene and whole genome sequencing [4,5], Luminex [6] and microarray analysis [7].

There is a wide range of technologies that provide specific short base sequences of DNA as probes — used to detect the complementary base sequence of interest—or as primers—that guide the DNA amplification process—used for different purposes. Primers and probes are the main components of nucleic acid-based detection systems and have been the subject of multiple studies. Therefore, different software programs have been developed to design these specific sequences of primers and probes minimizing potential cross-hybridization to be spotted, for example, as oligonucleotides in cDNA

* Correspondence: mgarcia@infomed.dia.fi.upm.es
† Contributed equally
Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegacedo S/N, 28660 Boadilla del Monte, Madrid, Spain
Full list of author information is available at the end of the article

Expanding the work on text mining to other Bioinformatics topics

Expanding research directions towards nano issues (UPM)

- Creating an inventory of nanobiomedical resources (towards a "**Nano Resourceome**")

- Text mining methods for extracting information from the literature and link them to augmented electronic health records (EHRs)

- A new approach for building nano-taxonomies based on quantitative and graphical information of nanoparticles

- Linking heterogeneous databases , including biomedical and nano-related information

The Action-Grid project is funded by the European Commission under the FP7

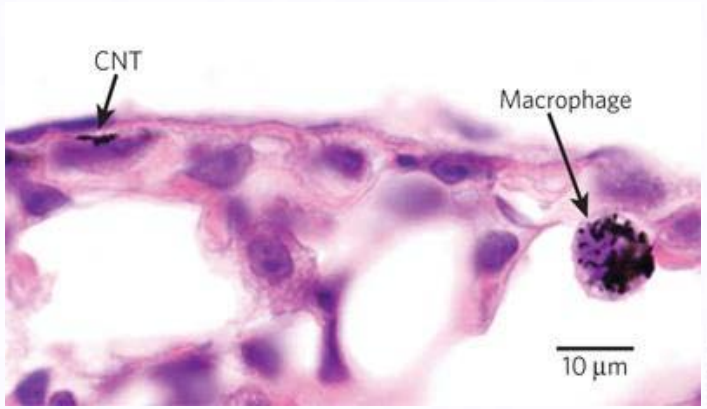
An Inventory of Nanoresources (primarily reported at KES 2008)

- Currently being developed at our group
- Provides detailed information on different nanoparticles/nanodevices/nanomaterials/tools
  - Morphology
  - Drug-delivery- related information
  - Nanotoxicology information
  - Informatics Tools
  - Relevant Publications
- All the information is automatically extracted from the literature using methods and tools borrowed from BMI

The Toxicity Searcher

- A literature retrieval system linking nanoparticles/nanomaterials to different anatomical locations
- Useful to retrieve papers dealing with drug delivery and nanotoxicity
- An application that relates ontologies

  from the BMI and Nanoinformatics fields



Research approach

- To create a system to retrieve documents about nanoparticles, their medical application, functionality and toxicity

- To build a literature retrieval search engine
- To create a system to generate and manage documental indexes
- To create a service to access information

Patient-based Biomedical Literature Retrieval

- Open source browser extension, a plug-in to connect nano-related information from to medical records
- EHR locally parsed using NLP techniques to identify relevant MeSH terms
- Automatic queries are generated regarding:
  - Pathology
  - EHR MeSH Terms
  - MeSH Qualifiers
  - Information about nanoparticles and secondary effects

BIOINFORSALUD - 2009

International Symposium on Research in Grid/Nano/Bio/Medical Informatics

ACTION-GRID

Main menu_

- home
- Invited Speakers
- Programme
- Organizing committee
- How to get there
- Contact

information_

BIOINFORSALUD 2009 will be held on **Monday 16th of March 2009** in "Palacio de Congresos de la Castellana", Paseo Castellana 99, Madrid. The symposium registration is FREE and it will be done in the site of the conference (online registration is NOT available).

links_

- ACTION-Grid Project
- EC-IST
- SEIS
- UPM
- ISCIII

BIOINFORSALUD 2009

BIOINFORSALUD 2009 is the International Symposium on Research in Grid/Nano/Bio/Medical Informatics.

It will be held on **Monday 16th of March 2009** in the **Palacio de Congresos de la Castellana - Madrid**

The symposium registration is FREE and it will be done in the site of the conference (online registration is NOT available).

A Nanoinformatics conference held in Madrid in 2009, part of ACTION Grid

European Commission
Information Society and Media

Victor Maojo